NOTE ON REPEATED SELECTION IN THE NORMAL CASE

BU-~~166~~ 171-M　　D. S. Robson　　April, 1964

ABSTRACT

A k-cycle selection model is specified by a (k+1)-variate normal distribution of the variables $X, Y_1 = X+\epsilon_1, \cdots, Y_k = X+\epsilon_k$ with selection at the i^{th} stage removing a fraction

$$P_i = P(Y_i > y_i \mid Y_1 > y_1, \cdots, Y_{i-1} > y_{i-1})$$

The distribution of X in this selected fraction is then convolved with the $N(0, \varepsilon_{i+1}^2)$ distribution of ϵ_{i+1} to form the distribution of Y_{i+1}. An expression is given for the characteristic function of X in the k^{th} selected fraction.

Biometrics Unit, Plant Breeding Department, Cornell University

NOTE ON REPEATED SELECTION IN THE NORMAL CASE

D. S. Robson

BU-~~166~~171-M April, 1964

Selection for a quantitative trait often continues for several cycles, as in the successive annual screening of a plant population in the process of developing new varieties. With plant selection, as with most other selection problems, the trait x being selected for cannot be measured without error, and actual selections are based on the observation $y_i = x + e_i$ in the i^{th} cycle of the process. We shall assume here that the error chance variable e_i is $N(0,\mathcal{E}_1^2)$ (normally distributed with mean 0 and variance $\mathcal{E}_1^2$) and that the error e_i attaching to x in the i^{th} stage is independent of the error e_j attaching to that same x (or any other x) in the j^{th} stage. Further, we suppose that in the unselected population the chance variable x is $N(\xi,\sigma^2)$, so that $y_1=x+e_1$ is $N(\xi,\omega_1^2 = \sigma^2 + \mathcal{E}_1^2)$.

The population available at the k^{th} stage is assumed to be of infinite size, and selection consists of removing the upper fraction P_k of the available y-population for further selection at stage k+1. The fraction of the original population available for selection at stage k+1 is therefore $P_1P_2\cdots P_k$, and our concern here shall lie with the distribution of x in this remaining fraction. These fractions are defined by

$$P_1 = P(Y_1 > y_1)$$

$$P_1P_2 = P_1P(Y_2 > y_2 \mid Y_1 > y_1)$$

$$P_1P_2\cdots P_k = P_1P_2\cdots P_{k-1}P(Y_k > y_k \mid Y_1 > y_1, Y_2 > y_2, \cdots, Y_{k-1} > y_{k-1})$$

and our results are based upon the observation that this remaining fraction is simply the tail probability in a k-variate normal distribution,

$$P_1P_2\cdots P_k = P(Y_1 > y_1, Y_2 > y_2, \cdots, Y_k > y_k)$$

Since the joint distribution of $X, Y_1, Y_2, \cdots, Y_k$ is the (k+1)-variate normal distribution with mean ξ and covariance matrix

$$\Lambda = \begin{bmatrix} \sigma^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \omega_1^2 & \cdots & \sigma^2 \\ \vdots & & & \\ \sigma^2 & \sigma^2 & \cdots & \omega_k^2 \end{bmatrix} = [\sigma_{ij}]$$

then the distribution of x for fixed values of $Y_1, \cdots, Y_k$ is normal with mean

$$E(X|y_1, \cdots, y_k) = \xi - \frac{\Lambda_{o1}}{\Lambda_{oo}}(y_1 - \xi) - \cdots - \frac{\Lambda_{ok}}{\Lambda_{oo}}(y_k - \xi)$$

and

$$\mathrm{var}(X|y_1, \cdots, y_k) = \frac{\Lambda}{\Lambda_{oo}}$$

where Λ is the determinant of Λ and Λ_{ij} is the cofactor of the ij^{th} element of Λ.

The joint distribution of $Y_1, \cdots, Y_k$ is normal with mean $\xi = (\xi, \cdots, \xi)$ and covariance matrix Λ_{oo}. Using the expansion

$$\Lambda = \sigma_{oo}\Lambda_{oo} - \sum_{i,j=1}^{k} \sigma_{io}\sigma_{oj}\Lambda_{oo\cdot ij}$$

where $\Lambda_{oo\cdot ij}$ is the cofactor of σ_{ij} in Λ_{oo}, we may then express the conditional moment generating function of X as

$$E(e^{tX}|y_1,\cdots,y_k)$$

$$= e^{t\xi+\frac{t^2}{2}(\sigma_{oo}-\frac{1}{\Lambda_{oo}}\sum_{i,j=1}^{k}\sigma_{io}\sigma_{oj}\Lambda_{oo\cdot ij})+\frac{t}{\Lambda_{oo}}\sum_{i,j=1}^{k}\sigma_{io}(y_j-\xi)\Lambda_{oo\cdot ij}}$$

and then

$$E(e^{tX}|Y_1>y_1,\cdots,Y_k>y_k) = e^{t\xi+\frac{t^2}{2}\sigma_{oo}}(P_1P_2\cdots P_k)^{-1}\frac{1}{(2\Pi)^{\frac{k}{2}}\sqrt{\Lambda_{oo}}} \cdot$$

$$\int_{\{u_i>y_i\}} e^{-\frac{1}{2\Lambda_{oo}}\sum_{i,j=1}^{k}[(u_i-\xi)(u_j-\xi)-2t\sigma_{io}(u_j-\xi)+t^2\sigma_{io}\sigma_{oj}]\Lambda_{oo\cdot ij}} du_1\cdots du_k$$

The exponent in the integral reduces to

$$\sum_{i,j=1}^{k}(u_i-\xi-\sigma_{io}t)(u_j-\xi-\sigma_{oj}t)\Lambda_{oo\cdot ij}$$

hence, transforming to the standard normal $z_i = (y_i-\xi)/\sqrt{\sigma_{ii}}$, we obtain

$$E(e^{tX}|\frac{Y_1-\xi}{\omega_1}>z_1,\cdots,\frac{Y_k-\xi}{\omega_k}>z_k)$$

$$= \frac{e^{t\xi+\frac{t^2}{2}\sigma^2}}{P_1P_2\cdots P_k}\frac{1}{(2\Pi)^{\frac{k}{2}}\sqrt{R_{oo}}}\int_{\{v_i>z_i-\frac{\sigma^2}{\omega_i}t\}} e^{-\frac{1}{2P_{oo}}\sum_{i,j=1}^{k}R_{oo\cdot ij}v_iv_j} dv_1\cdots dv_k$$

where

$$R_{00} = \begin{vmatrix} 1 & \frac{\sigma^2}{\omega_1\omega_2} & \cdots & \frac{\sigma^2}{\omega_1\omega_k} \\ \frac{\sigma^2}{\omega_1\omega_2} & 1 & \cdots & \frac{\sigma^2}{\omega_2\omega_k} \\ \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_k} & \frac{\sigma^2}{\omega_2\omega_k} & \cdots & 1 \end{vmatrix}$$

The mean value of X in this selected fraction of the population is obtained by differentiating once with respect to t, first writing

$$E(e^{tX} \mid \frac{Y_1-\xi}{\omega_1} > z_1, \cdots, \frac{Y_k-\xi}{\omega_k} > z_k)$$

$$= \varphi_X(t) P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t)/P_1P_2\cdots P_k$$

so that the derivative becomes

$$\frac{1}{P_1P_2\cdots P_k} \{\varphi_X'(t) P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t)$$

$$+ \frac{\sigma^2}{\sqrt{2\Pi}} \sum_{j=1}^{k} \frac{1}{\omega_j} e^{-\frac{1}{2}(z_j - \frac{\sigma^2}{\omega_j} t)^2}$$

$$P_{R_{00}}(v_1 > z_1 - \frac{\sigma^2}{\omega_1} t, \cdots, v_{j-1} > z_{j-1} - \frac{\sigma^2}{\omega_{j-1}} t, v_{j+1} > z_{j+1}$$

$$- \frac{\sigma^2}{\omega_{j+1}} t, \cdots, v_k > z_k - \frac{\sigma^2}{\omega_k} t | z_j)\}$$

Setting t=0, we obtain the mean value

$$\xi + \frac{\sigma^2}{P_1 P_2 \cdots P_k} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\Pi}} e^{-\frac{z_j^2}{2}} P_{R_{oo}}(v_1 > z_1, \cdots, v_{j-1} > z_{j-1}, v_{j+1} > z_{j+1}, \cdots,$$

$$v_k > z_k | v_j = z_j)$$

or

$$\xi + \frac{\sigma^2}{P_1 P_2 \cdots P_k} \sum_{j=1}^{k} \frac{1}{\omega_j \sqrt{2\Pi}} e^{-\frac{z_j^2}{2}} P_{R_{oo}^{(j)}} (u_1 > z_1 - \frac{\sigma^2}{\omega_1 \omega_j} z_j, \cdots, u_{j-1} > z_{j-1}$$

$$- \frac{\sigma^2}{\omega_{j-1}\omega_j} z_j, u_{j+1} > z_{j+1} - \frac{\sigma^2}{\omega_j \omega_{j+1}} z_j, \cdots, u_k > z_k - \frac{\sigma^2}{\omega_j \omega_k} z_j)$$

where

$$R_{00}^{(j)} = \begin{vmatrix} 1-\frac{\sigma^4}{\omega_1^2\omega_j^2} & \cdots & \frac{\sigma^2}{\omega_1\omega_{j-1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \frac{\sigma^2}{\omega_1\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_1\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_{j-1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & 1-\frac{\sigma^4}{\omega_{j-1}^2\omega_j^2} & \frac{\sigma^2}{\omega_{j-1}\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \frac{\sigma^2}{\omega_1\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_{j+1}}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & 1-\frac{\sigma^4}{\omega_j^2\omega_{j+1}^2} & \cdots & \frac{\sigma^2}{\omega_{j+1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\sigma^2}{\omega_1\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & \frac{\sigma^2}{\omega_{j-1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \frac{\sigma^2}{\omega_{j+1}\omega_k}\left(1-\frac{\sigma^2}{\omega_j^2}\right) & \cdots & 1-\frac{\sigma^4}{\omega_j^2\omega_k^2} \end{vmatrix}$$